

A Method and System for Coherently Caching I/O Devices Across a Network

Abstract

The cache keeps regularly accessed disk I/O data within RAM that forms part of a computer systems main memory. The cache operates across a network of computers systems, maintaining cache coherency for the disk I/O devices that are shared by the multiple computer systems within that network. Read access for disk I/O data that is contained within the RAM is returned much faster than would occur if the disk I/O device was accessed directly. The data is held in one of three areas of the RAM for the cache, dependent on the size of the I/O access. The total RAM containing the three areas for the cache does not occupy a fixed amount of a computers main memory. The RAM for the cache grows to contain more disk I/O data on demand and shrinks when more of the main memory is required by the computer system for other uses. The user of the cache is allowed to specify which size of I/O access is allocated to the three areas for the RAM, along with a limit for the total amount of main memory that will be used by the cache at any one time.